AD-A098 074 PITTSBURGH UNIV PA INST FOR STATISTICS AND APPLICATIONS F/G 12/1
NEARLY EFFICIENT ESTIMATORS BASED ON ORDER STATISTICS.(U)
FEB 81 T KARIYA, B K SINHA, K SUBRAMANYAM F49620-79-C-0161
UNCLASSIFIED TR-81-05 AFOSR-TR-81-0388 NL

1 OF 1
AD A
098074

END
DATE
FILMED
5-81
DTIC

NEARLY EFFICIENT ESTIMATORS BASED ON ORDER STATISTICS*

Takeaki Kariya[1]
Bimal K. Sinha[2]
Kasala Subramanyam

University of Pittsburgh

February 1981

Technical Report No. 81-05

Institute For Statistics and Applications
Department of Mathematics and Statistics
University of Pittsburgh
Pittsburgh, PA 15260

*The work of Sinha and Subramanyam and part of the work of Kariya is sponsored by the Air Force office of Scientific Research under Contract F49620-79-C-0161. Part of the work of Kariya was done under Mellon Postdoctoral Fellowship program.

[1]On leave from Hitotsubashi University

[2]*On leave from the Indian Statistical Institute*

DISCLAIMER NOTICE

THIS DOCUMENT IS BEST QUALITY PRACTICABLE. THE COPY FURNISHED TO DTIC CONTAINED A SIGNIFICANT NUMBER OF PAGES WHICH DO NOT REPRODUCE LEGIBLY.

UNCLASSIFIED

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: AFOSR-TR- 81-0388
2. GOVT ACCESSION NO.: AD-A098074
3. RECIPIENT'S CATALOG NUMBER:
4. TITLE (and Subtitle): Nearly Efficient Estimators Based on Order Statistics
5. TYPE OF REPORT & PERIOD COVERED: INTERIM
6. PERFORMING ORG. REPORT NUMBER: U. P. Tech. Rept. 81-05
7. AUTHOR(s): Takeaki Kariya, Bimal K. Sinha and Kasala Subramanyam
8. CONTRACT OR GRANT NUMBER(s): F49620-79-C-0161
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Dept. of Mathematics & Statistics, University of Pittsburgh, Pittsburgh, PA 15260
10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: 2304/A5 61102F
11. CONTROLLING OFFICE NAME AND ADDRESS: AFOSR/NM, Bolling AFB, DC 20332
12. REPORT DATE: February 1981
13. NUMBER OF PAGES: 9
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office):
15. SECURITY CLASS. (of this report): UNCLASSIFIED
15a. DECLASSIFICATION/DOWNGRADING SCHEDULE:
16. DISTRIBUTION STATEMENT (of this Report): Approved for public release; distribution unlimited.
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report):
18. SUPPLEMENTARY NOTES:
19. KEY WORDS (Continue on reverse side if necessary and identify by block number): order statistics, location parameter, relative asymptotic efficiency.
20. ABSTRACT (Continue on reverse side if necessary and identify by block number):

In this note, based on 3 or 5 order statistics out of n order statistics, nearly asymptotically efficient estimators are obtained for Cauchy and logistic distributions as well as normal distributions.

DD FORM 1 JAN 73 1473

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE

81 4 22 056

ABSTRACT

In this note, based on 3 or 5 order statistics out of n order statistics, nearly asymptotically efficient estimators are obtained for Cauchy and logistic distributions as well as normal distributions.

1. INTRODUCTION

In 1946, to estimate the location parameter θ of normal distribution $N(\theta,1)$, Mosteller considered estimators of the form

$$\hat{\theta} = \sum_{i=1}^{k} c_i X_{(n_i)}, \tag{1.1}$$

where $X_{(1)} \leq \cdots \leq X_{(n)}$ are the order statistics of a random sample $X_1,\ldots,X_n$ from $N(\theta,1)$, $n_i = [n\alpha_i] + 1$ with $0 < \alpha_1 < \ldots < \alpha_k < 1$, c_i's are constants and $[n\alpha_i]$ is the largest integer less than or equal to $n\alpha_i$. Assuming equal weights $c_1 = \ldots = c_k = 1/k$, for each $k = 1,2,\ldots$, he found the optimal choice for $(\alpha_1,\ldots,\alpha_k)$ for which the asymptotic variance of $\hat{\theta}$ is minimized. In this note, we generalize his work into two directions. First, rather than normal distribution, an arbitrary distribution of a location parameter θ with pdf $f(x-\theta)$ is considered, where f is assumed to be known. Second, instead of giving equal weights for $(c_1,\ldots,c_k)$, the optimum weights are determined along with the optimum spacings for $(n_1,\ldots,n_k)$ or $(\alpha_1,\ldots\alpha_k)$. Although our argument is applicable to general $k = 1,2,\ldots$, because of analytical complication and because of satisfactory achievement of relative asymptotic efficiency, we discuss the case of $k = 3$ in details and briefly treat the case of $k = 5$. In section 3, the result is applied to the Cauchy distribution with pdf

$$f(x-\theta) = 1/\pi[1+(x-\theta)^2], \quad -\infty < \theta < \infty \tag{1.2}$$

and the logistic distribution with pdf

$$f(x-\theta) = e^{-(x-\theta)}/[1+e^{-(x-\theta)}]^2, \quad -\infty < \theta < \infty \tag{1.3}$$

as well as normal distribution. As is well known, a computational difficulty is involved in deriving the MLE's (maximum likelihood estimators) of θ for the distributions (1.2) and (1.3). On the other hand, it is also well known (e.g.[11]) that for any $f(x-\theta)$ satisfying certain regularity conditions, with $k = n$ and suitable $c_i = c_i(n,f)$ (known as score function), $\hat{\theta}$ in (1.1) becomes a BAN (best asymptotically normal) estimator. Compared to the BAN estimators, our estimators have a great computational merit and yet are nearly efficient. In fact, the estimators in this note use only 3 or 5 order statistics out of n order statistics however large n may be and attain at least 90% asymptotic efficiency, relative to the BAN estimators. More specifically, in the Cauchy case the RAE (relative asymptotic efficiency) of the median alone is 81%, but the RAE's of our optimum estimators with $k = 3$ and $k = 5$ are 87% and 90% respectively. In the logistic case, while the RAE of the median alone is only 75%, the RAE of our optimum estimator with $k = 3$ is surprisingly 93.8%. It is noted that the optimal weights c_i's in these estimators are constants independently of n. Further, our estimators are briefly compared to similar estimators considered in Gastwirth (1966).

We remark that, associated with robust estimators, estimators of the form (1.1) with $k = n$ have been treated in great many papers (e.g. [5]).

2. <u>Main result.</u> Let F be the class of continuously differentiable and symmetric pdf's on R^1 with respect to the Lebesgue measure, and suppose $X_1,\ldots,X_n$ are i.i.d. with pdf $f(x-\theta)$ where $f \in F$ is known and let $X_{(1)} \leq \cdots \leq X_{(n)}$ be the corresponding order statistics. Further, for any k, let $0 < \alpha_1 < \ldots < \alpha_k < 1$ be real numbers and let ξ_i be the α_i - quantile of the population when $\theta = 0$, i.e.,

$$\alpha_i = \int_{-\infty}^{\xi_i} f(x)dx = F(\xi_i) \qquad (2.1)$$

We assume that $f(\xi_i) > 0$ $(i=1,\ldots,k)$. Define the sample quantiles by $X_{(n_i)}$'s where $n_i = [n\alpha_i] + 1$ $(i=1,\ldots,k)$. To estimate the location parameter θ, we consider the class of linear estimators of $X_{(n_i)}$'s of the form,

$$\hat{\theta} = a_1 X_{(n_1)} + \ldots + a_k X_{(n_k)}, \quad \text{where} \quad (a_1,\ldots,a_k) \in R^k. \qquad (2.2)$$

As is well known ([4] p. 17), as $n \to \infty$, the joint distribution of $(X_{(n1)},\ldots,X_{(n_k)})$ converges to a k-variate normal distribution with mean $(\theta + \xi_1,\ldots,\theta + \xi_k)$ and covariance matrix $\frac{1}{n}((\sigma_{ij}))$, where

$$\sigma_{ij} = \alpha_i(1-\alpha_j)/f(\xi_i)f(\xi_j) \quad (i \leq j).$$

Therefore, for any $\hat{\theta}$ in (2.2), the asymptotic distribution of $\hat{\theta}$ is normal with mean $\mu = \sum_{i=1}^{k} a_i(\theta + \xi_i)$ and variance σ^2/n, where

$$\sigma^2 = \sum_{i=1}^{k} a_i^2 \alpha_i(1-\alpha_i)/f^2(\xi_i) + \sum_{i<j} a_i a_j \alpha_i (1-\alpha_j)/f(\xi_i)f(\xi_j)$$
$$+ \sum_{i>j} a_i a_j \alpha_j (1-\alpha_i)/f(\xi_i)f(\xi_j) \qquad (2.4)$$

For $\hat{\theta}$ to be consistent for θ, the conditions

$$\sum_{i=1}^{k} a_i = 1 \quad \text{and} \quad \sum_{i=1}^{k} a_i \xi_i = 0 \qquad (2.5)$$

are necessary ans sufficient. Let C_k be the class of estimators of the form (2.2) satisfying (2.5). Then finding a minimum asymptotic variance estimator in C_k is equivalent to minimizing (2.4) with respect to $(a_1,\ldots a_k)$ and $(\alpha_1,\ldots,\alpha_k)$ subject to (2.5). Due to the symmetry of f, for k odd, we include the median $X_{([\frac{n}{2}]+1)}$ and choose the other quantiles symmetrically about the median, i.e., $\alpha_1 = 1-\alpha_k$, $\alpha_2 = 1-\alpha_{k-1}, \ldots$ or equivalently $\xi_1 = -\xi_k$, $\xi_2 = -\xi_{k-1}, \ldots$, and the weights $a_1 = a_k$, $a_2 = a_{k-1}, \ldots$. Then, clearly the second condition of (2.5) holds and $f(\xi_1) = f(\xi_k)$, $f(\xi_2) = f(\xi_{k-2}), \ldots$.

Case k = 3. In this case, $\hat{\theta}$ in (2.2) is of the form

$$\hat{\theta}(a,b,\alpha_1) = aX_{(1)} + bX_{(2)} + aX_{(3)}$$

where $X_{(1)}$ and $X_{(3)}$ are the sample α_1 and $(1-\alpha_1)$ quantiles respectively and $X_{(2)}$ is the median. From the first condition of (2.5), $2a+b=1$. Writing $f(\xi_1) = f_1$ and $f(\xi_2) = f(0) = f_o$, the asymptotic variance of $\hat{\theta}$ in (2.4) is computed as

$$\sigma^2(a,\alpha_1) = a^2\left[\frac{2\alpha_1}{f_1^2} - \frac{4\alpha_1}{f_1 f_o} + \frac{1}{f_o^2}\right] + a\left[\frac{2\alpha_1}{f_1 f_o} - \frac{1}{f_o^2}\right] + \frac{1}{4f_o^2} . \qquad (2.7)$$

Minimizing this with respect to a yields

$$a_* = a_*(\alpha_1) = \left[\frac{1}{f_o^2} - \frac{2\alpha_1}{f_1 f_o}\right]/2\left[\frac{2\alpha_1}{f_1^2} - \frac{4\alpha_1}{f_1 f_o} + \frac{1}{f_o^2}\right] \quad \text{and} \qquad (2.8)$$

$$\sigma^2(a_*,\alpha_1) = \alpha_1(1-2\alpha_1)/2[2\alpha_1 f_o^2 - 4\alpha_1 f_1 f_o + f_1^2] . \qquad (2.9)$$

To further minimize (2.9) with respect to α_1 $(0 < \alpha_1 < 1/2)$, let

$$\eta(\alpha_1) = f_1/f_o = f(F^{-1}(\alpha_1))/f(0), \qquad (2.10)$$

where F^{-1} is the inverse of F in (2.7). Then minimizing (2.9) is equivalent to maximizing

$$1/f_o^2\,\sigma^2(a_*,\alpha_1) = (\eta-2\alpha_1)^2/\alpha_1(1-2\alpha_1) + 2, \tag{2.10}$$

which results in the equation

$$\eta'(\alpha_1) - [(1-4\alpha_1)/2\alpha_1(1-2\alpha_1)]\,\eta(\alpha_1) = 1/(1-2\alpha_1) \tag{2.11}$$

In any specific problem, we first solve (2.11) for $\alpha_1 (o < \alpha_1 < 1/2)$, and then compute a_* via (2.8). Finally from (2.9) the minimum asymptotic variance $\sigma_*^2 \equiv \sigma^2(a_*(\alpha_{1*}), \alpha_{1*})$ is obtained, where α_{1*} is the solution of (2.11).

As a remark, in order that σ_*^2 be less than the asymptotic variance of the median $X_{(2)}$, i.e., $1/4f_o^2$, it is easy to see that

$$(f_{1*}-2f_o\alpha_{1*})^2 > o \tag{2.12}$$

is necessary and sufficient. Hence, unless

$$f(F^{-1}(\alpha_{1*})) = 2f_o\alpha_{1*}, \tag{2.13}$$

$\hat\theta_* = \hat\theta(a_*, b_*, \alpha_{1*})$ is asymptotically better than the median $X_{(2)}$. Since (2.13) is implied by

$$f(x) = 2f(0)F(x) \quad \text{for all } x, \tag{2.14}$$

it follows in particular that in the case of the double exponential density $f(x) = \exp(-|x|)/2$, $\hat\theta_*$ cannot improve over $X_{(2)}$. Further, we remark that one of the optimal weights (a^*,b^*) may take negative value.

<u>Case k=5.</u> In this case, $\hat\theta$ in (2.2) is of the form

$$\hat\theta = \hat\theta(a,b,c,\alpha_1,\alpha_2) = aX_{(1)} + bX_{(2)} + cX_{(3)} + bX_{(4)} + aX_{(5)}, \tag{2.15}$$

where $X_{(1)}$, $X_{(2)}$, $X_{(4)}$ and $X_{(5)}$ are respectively the α_1, α_2, $(1-\alpha_2)$, and $(1-\alpha_1)$ quantiles, and $X_{(3)}$ is the median $(o < \alpha_1 < \alpha_2 < 1/2)$. The first condition of (2.5) yields $2a + 2b + c = 1$. Let $f_1 = f(\xi_1)$, $f_2(\xi_2) = f_2$ and $f(0) = f_o$. Then σ^2 in (2.4) reduces to

$$\sigma^2(a,b,c,\alpha_1,\alpha_2) = a^2A + b^2C + abB + aD + bE + 1/4f_o^2$$

where

$$\left.\begin{aligned} A &= \frac{2\alpha_1}{f_1^2} - \frac{4\alpha_1}{f_1 f_o} + \frac{1}{f_o^2}, \qquad B = \frac{4\alpha_1}{f_1 f_2} - \frac{4\alpha_1}{f_1 f_o} - \frac{4\alpha_2}{f_2 f_o} + \frac{2}{f_o^2} \\ C &= \frac{2\alpha_2}{f_2^2} - \frac{4\alpha_2}{f_2 f_o} + \frac{1}{f_o^2}, \qquad D = \frac{2\alpha_1}{f_1 f_o} - \frac{1}{f_o^2} \\ E &= \frac{2\alpha_2}{f_2 f_o} - \frac{1}{f_o^2} \end{aligned}\right\} \quad (2.17)$$

Minimizing (2.16) with respect to (a,b) yields

$$\left.\begin{aligned} a_* &= (BE - 2CD)/(4AC - B^2) \\ b_* &= (2AE - BD)/(B^2 - 4AC) \end{aligned}\right\} \quad (2.18)$$

and

$$\sigma^2(a_*\ b_*,\ c_*,\ \alpha_1,\ \alpha_2) = \frac{CD^2 + BDE - AE^2}{B^2 - 4AC} + \frac{1}{4f_o^2}. \qquad (2.19)$$

For ready reference, we record below

$$B^2-4AC = \frac{(4\alpha_1-1)^2}{f_1^2}\left(\frac{1}{f_2} - \frac{1}{f_o}\right)^2 \quad \left(\frac{4\alpha_2-1}{f_2 f_o}\right)^2 + \frac{16\alpha_1}{f_1 f_2 f_o}\left(\frac{1}{f_o} - \frac{1}{f_1}\right)$$

$$+ \frac{8\alpha_1}{f_1^2 f_2^2} - \frac{1}{f_2^2 f_o^2} + \frac{16\alpha_1\alpha_2}{f_1 f_2}\left(\frac{2}{f_1 f_o} - \frac{1}{f_1 f_2} - \frac{2}{f_o^2}\right) - \frac{1}{f_1^2}\left(\frac{1}{f_2} - \frac{1}{f_1}\right)^2$$

and

$$CD^2 + BDE - 2AE^2 = \left(\frac{2\alpha_1}{f_1 f_0} - \frac{1}{f_0}\right)^2 \left[\frac{2\alpha_1}{f_0^2} - \frac{4\alpha_2}{f_2 f_0} + \frac{1}{f_0^2}\right] \tag{2.21}$$

$$+\left(\frac{2\alpha_2}{f_2 f_0} - \frac{1}{f_0^2}\right)^2 \left[\frac{6\alpha_1}{f_1 f_0} - \frac{1}{f_0^2} - \frac{6\alpha_1}{f_1^2}\right] - \frac{4\alpha_1}{f_1}\left(\frac{1}{f_0} - \frac{1}{f_2}\right)\left(\frac{2\alpha_2}{f_2 f_0} - \frac{1}{f_0^2}\right)\left(\frac{2\alpha_1}{f_1 f_0} - \frac{1}{f_0^2}\right).$$

Exact algebraic minimization of $\sigma^2(a_*, b_*, c_*, \alpha_1, \alpha_2)$ with respect to α_1 and α_2 seems difficult although in specific problems, the solution can be obtained numerically. It is remarked that even an approximate solution of (α_1, α_2) coupled with the optimal weights (a^*, b^*) in (2.18) may be useful in getting $\hat{\theta}$ better than the median and closer to Chernoff, Gastwirth and Johns' BAN estimator (see [1]).

3. Comparison of estimators. The table 1 below summarizes features of our optimum estimators for $k = 3$ for normal, Cauchy and logistic distributions.

TABLE 1 ($k = 3$)

	α_{1*}	a_*	$b^*=1-2a^*$	σ_*^2	$^1/I$	ARE
Normal	0.1632	0.3054	0.3892	1.1100	1	90%
	(1/3)	(0.3)	(0.4)	(1.2820)		(78%)
Cauchy	0.0791	-0.0207	1.0414	2.2023	2	86.9%
	(1/3)	(0.3)	(0.4)	(2.5000)		(80%)
Logistic	0.2500	0.3	0.4	3.2	3	93.8%
	(1/3)	(0.3)	(0.4)	(3.2867)		(91%)

Here the numbers in the parentheses are those obtained from the viewpoint of robustness by Gastwirth (1966), and 1/I is the asymptotic minimum variance of the BAN estimator where I denotes the Fisher information. In the normal case, while the ARE of Mosteller's estimator with $k = 3$, $\alpha_1 = 0.1826$ and equal weights is 88%, the ARE of our estimator is 90%. Compared to Gastwirth's robust estimator, which is the same form as our estimators, naturally our estimators perform better since in our case, knowing the form of f, the optimal weights with optimal spacings are obtained. This might suggest that, when n is large, it will be better to try to know the form of f rather than to apply "robust" procedures. However, for example, testing the normality in an efficient way is not an easy task. Since for $k = 3$, the RAE of Cauchy case is still less than 90%, we computed the case of $k = 5$ numerically. The resulting optimum values are as follows

$$a_* = -0.0720,\ b_* = 0.0435,\ c_* = 0.0312$$

$$\alpha_{1*} = 0.16,\ \alpha_{2*} = 0.19 \text{ and } \sigma_*^2 = 2.22 .$$

In this case, the RAE of $\hat{\theta}_*$ is 90.1%. It is noted that when $k = 3$ and $k = 5$, some of the optimal weights are negative, which is consistent with the result by Chernoff, Gastwirth and Johns (1967).

Acknowledgement: The authors thank Mr. Robert Boudreau for his help in the computations.

REFERENCES

[1] Chernoff, H., Gastwirth, J. L., and Johns, M. V. (1967). Asymptotic distribution of linear combinations of functions of order statistics with applications to estimator. Ann. Math. Statist. 38 52-72.

[2] Gastwirth, J. L. (1966). On robust procedures. Journal of the Amer. Statist. Assoc. 61 929-948.

[3] Mosteller, F. (1946). On some useful 'inefficient' statistics. Ann. Math. Statist. 17 377-408.

[4] Sarhan, A. E. and Greenberg, B. G. (1962). Contributions to Order Statistics. John Wiley and Sons, New York.

[5] Tukey, J. W., Bickel, P. J., et al. (1972). Robust Estimates of Locations. Princeton University Press.

END

DATE

FILMED

5-81

DTIC